STATISTICS IN MEDICINE
Statist. Med. 2000; **19**:2493–2519

Statistical issues in the analysis of disease mapping data

C. Pascutto[1,2], J. C. Wakefield[2,*,†], N. G. Best[2], S. Richardson[3], L. Bernardinelli[1], A. Staines[4] and P. Elliott[2]

[1]*Dipartimento di Scienze Sanitarie Applicate e Psicocomportamentali, Universitá di Pavia, Italy*
[2]*Department of Epidemiology & Public Health, Imperial College School of Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, U.K.*
[3]*INSERM Unit 170, 16 av Paul-Vaillant Couturier, 94807 Villejuif Cedex, France*
[4]*Department of Public Health Medicine and Epidemiology, UCD, Earlsfort Terrace, Dublin 2, Ireland*

SUMMARY

In this paper we discuss a number of issues that are pertinent to the analysis of disease mapping data. As an illustrative example we consider the mapping of larynx cancer across electoral wards in the North West Thames region of the U.K. Bayesian hierarchical models are now frequently employed to carry out such mapping. In a typical situation, a three-stage hierarchical model is specified in which the data are modelled as a function of area-specific relative risks at stage one; the collection of relative risks across the study region are modelled at stage two; and at stage three prior distributions are assigned to parameters of the stage two distribution. Such models allow area-specific disease relative risks to be 'smoothed' towards global and/or local mean levels across the study region. However, these models contain many structural and functional assumptions at different levels of the hierarchy; we aim to discuss some of these assumptions and illustrate their sensitivity. When relative risks are the endpoint of interest, it is common practice to assume that, for each of the age–sex strata of a particular area, there is a common multiplier (the relative risk) acting upon each of the stratum-specific risks in that area; we will examine this proportionality assumption. We also consider the choices of models and priors at stages two and three of the hierarchy, the effect of outlying areas, and an assessment of the level of smoothing that is being carried out. For inference, we concentrate on the description of the spatial variability in relative risks and on the association between the relative risks of larynx cancer and an area-level measure of socio-economic status. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Disease mapping may be carried out for a variety of reasons including simple description, and the assessment of the degree of spatial and non-spatial variability in rates/relative risks.

*Correspondence to: J. C. Wakefield, Department of Epidemiology and Public Health, Imperial College School of Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, U.K.
† E-mail: j.c.wakefield@ic.ac.uk

Contract/grant sponsor: Pan Thames Environmental R&D Programme; contract/grant number: 339
Contract/grant sponsor: ESRC; contract/grant number: H519255036
Contract/grant sponsor: EU BIOMED II; contract/grant number: PL96 3488
Contract/grant sponsor: The Wellcome Trust; contract/grant number: 0455051/Z/95/Z

A model-based approach may also be used to determine the ecological association between area-level explanatory variables and disease risk [1]. In this paper we examine a number of issues related to the estimation and interpretation of disease mapping data. We do this in the context of a study of larynx cancer incidence in 213 electoral wards in the North West Thames region of the United Kingdom (U.K.) in the period 1985–1993. We initially consider data by sex and by five age-bands: 35–44; 45–54; 55–64; 65–74, and 75 years and over. There are very few cases in those below 35 years of age and so we do not consider this group.

1.1. Epidemiological background

Laryngeal cancer is relatively common in males, though much less common in females. The male : female ratio is typically between 6:1 and 10:1. In France, the male age standardized incidence rate varies from 6.0/10 000 per year to 14.8/10 000 per year between different cancer registries. In Britain, the range is from 3.2 to 6.9, with the highest rates in Scotland [2]. The incidence of laryngeal cancer is rising in many European countries, falling in some, and remaining steady in others [3]. While some of this variation is due to differences and changes in smoking behaviour, other causes, including changes in alcohol intake, are probably operating also. Trends in mortality do not follow trends in incidence closely, perhaps due to improving treatment [3]. Little is known about smaller scale spatial variation in the incidence of laryngeal cancer.

Laryngeal cancer is mainly a disease of smokers, and both relative and attributable risks of laryngeal cancer amongst smokers are very high [4–6]. Alcohol use is a very important risk factor in most [5, 7, 8], but not all, studies [9]. There is evidence of a link with poverty [10] and poor diet is also an important risk factor; studies have shown increased risks associated with diets poor in vitamins C and E and beta carotene, and with high intakes of preserved meat, animal fats and with low intakes of fruit and dark green vegetables [6, 9, 11, 12]. Occupational risks have been identified, but their importance remains unclear [7, 9, 13]. Two studies of reported clusters of laryngeal cancer around industrial sites have produced negative results [14, 15]. The relationship between laryngeal papillomas, caused by a viral infection, and laryngeal cancer is still obscure [16].

1.2. Data

The raw data are in the form of disease counts, Y_{ij}, and population counts, N_{ij}, where $i = 1, \ldots, I$, indexes areas (wards) and $j = 1, \ldots, J$ indexes strata (age–sex groups). For the larynx cancer data we have $I = 213$ and (initially) $J = 10$. For rare and non-infectious diseases we may then assume

$$Y_{ij} \sim \text{Poisson}(N_{ij}\, p_{ij}) \tag{1}$$

where p_{ij} is the incidence of larynx cancer in ward i and strata j. Note that this model assumes constant risk within each area $\times$ stratum combination. In many cases this is a reasonable working hypothesis, although the discretization into areas and strata is an artificial construct and does not necessarily coincide with changes in risk. Further discussion of the tenability of the Poisson assumption may be found in Hoem [17]. Errors in the numerator (for example, under-registration of cases), and the denominators (for example, under-enumeration at census) may also lead to inadequacies in the Poisson model; errors of this type are described more extensively in Wakefield and Elliott [18].

In model (1) we have $I \times J$ probabilites to estimate and it is clear that, in general, a reduction in the number of parameters is required. For example, for the larynx cancer data, there are 2130

Table I. Summary of health, population and deprivation data across 213 Thames wards for the larynx cancer data.

	Minimum	25%	50%	75%	Maximum	Mean
Cases, male	0	0	3	4	12	2.8
Cases, female	0	0	0	1	3	0.5
Cases, total	0	2	3	4	12	3.3
Expecteds, male	1.1	2.1	2.7	3.4	5.7	2.8
Expecteds, female	0.2	0.4	0.5	0.6	1.1	0.5
Expecteds, total	1.3	2.4	3.2	4.0	6.8	3.3
Populations, male ($\times 10^3$)	0.9	1.6	2.0	2.4	4.1	2.0
Populations, female ($\times 10^3$)	1.0	1.7	2.2	2.8	5.0	2.3
Populations, total ($\times 10^3$)	2.0	3.3	4.2	5.3	9.1	4.4
SMRs, male	0	0.5	0.9	1.5	7.8	1.1
SMRs, female	0	0	0	2.3	8.2	1.1
SMRs, total	0	0.5	0.9	1.4	6.6	1.1
Carstairs deprivation index	−3.6	−0.4	1.6	3.9	11.7	2.0

probabilities in the saturated model with just 703 cases. Therefore, the model is usually simplified by assuming that

$$p_{ij} = \theta_i \times q_j \tag{2}$$

where θ_i represents the *relative risk* in area i and q_j is the *reference* incidence in stratum j. Under this assumption we have

$$Y_i \sim \text{Poisson}(E_i\theta_i) \tag{3}$$

where $Y_i = \sum_j Y_{ij}$ and $E_i = \sum_j N_{ij}q_j$ denote the observed and *expected* number of cases in area i.

Table I displays summary statistics for the larynx cancer data. We note that, as expected, there are far fewer cases for women than for men. For each sex the expected numbers are calculated over the five age groups with $q_j = \sum_i Y_{ij} / \sum_i N_{ij}$, $j = 1, \ldots, 5$. We note that standardization in this way will produce reference age probabilities that will also reflect environmental exposures if these are not distributed independently across age groups. The table also summarizes the Carstairs measure of socio-economic deprivation [19]. This index is calculated from census variables concerning overcrowding, access to a car, social class of head of household and unemployment. These variables are standardized and combined to give a continuous score, with high values indicating greater deprivation. Using this index allows some account of socio-economic status to be taken. Socio-economic status is well known to be a strong predictor of disease [20]. The mean of this index across the U.K. as a whole is zero and so we see that the study region is relatively deprived (though there is a large spread of deprivation across the region). From this point onwards we consider males only due to the sparsity of female cases.

The analyses that we present in Sections 3 and 4 are intended for two purposes: for description in order to examine the spatial variability in larynx cancer relative risks, and for consideration of the ecological association between the risk of larynx cancer and the Carstairs measure of socio-economic status.

The structure of this paper is as follows. In Section 2 we describe a three-stage hierarchical model that has been conventionally used for disease mapping and highlight key assumptions within this model. In Section 3 we present initial exploratory analyses of the larynx cancer data, including an examination of the proportionality assumption (2). In Section 4 we present a more comprehensive analysis and in Section 5 provide a concluding discussion.

2. STATISTICAL FRAMEWORK

In this section we describe a three-stage hierarchical model which may be used to analyse disease mapping data; Mollié [21] contains further details. We begin by assuming that the first stage model given in equation (3) is appropriate, and return to the assessment of proportionality in the next section. The maximum likelihood estimator (MLE) of the relative risks from model (3) corresponds to a standardized mortality/morbidity ratio (SMR)

$$\hat{\theta}_i = \frac{Y_i}{E_i} \tag{4}$$

The variance of this estimator is proportional to E_i^{-1} and so for areas with small populations there will be high sampling variability. Thus, for example, if the SMRs are mapped, large rural areas with low populations will often appear to display high risk due to the high variability of these estimates. The mapping of significance levels in order to overcome this problem produces its own difficulties since areas with large populations are more likely to attain significance, even if the excess risk is small [22]. Another difficulty with the use of SMRs for inference is that, for small areas in particular, SMRs in areas that are geographically close tend to display positive dependence, that is, positive spatial autocorrelation. If this aspect is ignored, incorrect inference will result; in particular, standard errors of ecological regression coefficients will be too small. Figure 1 shows the mapped SMRs for larynx cancer for males in the Thames study area. From this map and Table I we see that there is a reasonable amount of variability in relative risks across wards, but interpretation is difficult due to the aforementioned problems.

To overcome these problems, Besag *et al.* [23] suggested combining (3) with the following model for the relative risks:

$$\log \theta_i = X_i^{\mathrm{T}}\beta + V_i + U_i \tag{5}$$

where $X_i = (X_{i1}, \ldots, X_{ik})^{\mathrm{T}}$ is a $k \times 1$ vector of area-level risk factors, $\beta = (\beta_1, \ldots, \beta_k)^{\mathrm{T}}$ is a $k \times 1$ vector of regression parameters (with e^{β_l} representing the relative risk due to risk factor X_l, $l = 1, \ldots, k$), V_i, $i = 1, \ldots, I$ represents a 'residual' with no spatial structure (so that V_i and V_j are independent for $i \neq j$), and U_i, $i = 1, \ldots, I$ represents a 'residual' with spatial structure (so that U_i and U_j are modelled to have positive spatial dependence).

Model (5) forms the second stage of the hierarchical model. We let $U = (U_1, \ldots, U_I)^{\mathrm{T}}$, $V = (V_1, \ldots, V_I)^{\mathrm{T}}$, and ψ denote the hyperparameters, that is, the parameters of the variance-covariance matrices of the distributions of U and V. As discussed above, unmeasured risk factors, artificial areal-stratum boundaries and data inaccuracies may lead to the Poisson model (3) being inadequate. In particular we typically find that $\mathrm{var}(Y) > E(Y)$, that is, the area-specific disease counts exhibit overdispersion. Model (5) allows the spatial and non-spatial modelling of this overdispersion via the random effects U_i and V_i, respectively. If we believed that non-spatial overdispersion only

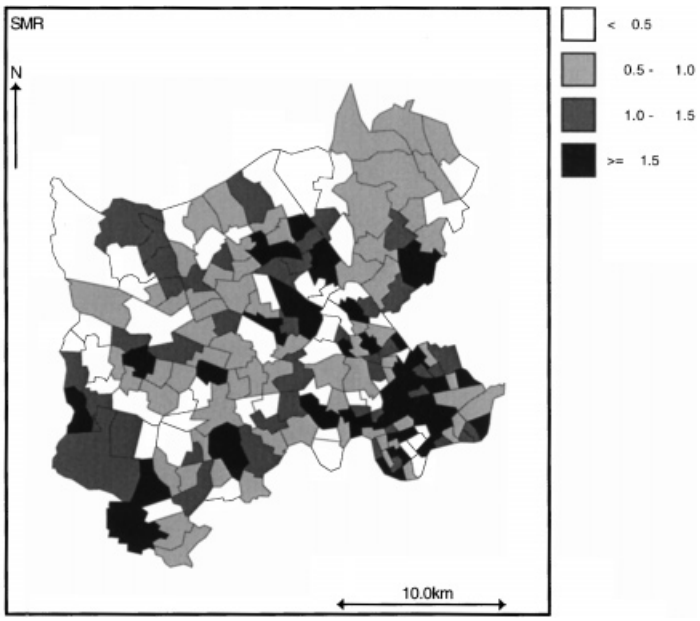

Figure 1. Larynx cancer SMRs for men, for 213 wards in the Thames region over the period 1985–1993.

needed to be considered, then it is natural to model $\exp(V_i)$ as arising from a gamma distribution as this leads to a tractable marginal distribution (the negative binomial). The gamma distribution cannot easily be extended to allow positive spatial dependence (but see Wolpert and Ickstadt [24] for recent developments). However, the normal distribution does allow such an extension and is the common choice for both spatial and non-spatial random effects. For the unstructured random effects it is therefore often assumed that $V_i \sim_{\text{i.i.d.}} \text{N}(0, \sigma_v^2)$.

For the spatially-dependent random effects in (5), the problem is to model the I-dimensional random variable U allowing for dependence between U_i and U_j, $i \neq j$. At this stage, the modelling choice is speculative and, contrary to other spatial analyses, little guided by scientific consideration. We may proceed either by trying to specify the *joint distribution* of U, or by using the univariate *conditional distributions* $U_i|U_j = u_j,\ j \neq i,\ i = 1, \ldots, I$.

In the joint modelling approach a common model for U is the zero mean multivariate normal distribution $\text{N}_I(0, \sigma_u^2 \Sigma(\phi))$. The $I \times I$ positive definite correlation matrix $\Sigma(\phi)$ contains elements $\Sigma_{ij}(\phi)$, $i, j = 1, \ldots, I$ with diagonal elements equal to one and off-diagonal elements describing the correlation between U_i and U_j, $i \neq j$; ϕ is a c-dimensional vector of unknown parameters. Various structured forms may be assumed for $\Sigma(\phi)$. A common choice is to assume that the dependence is a function of the distance, d_{ij}, between the population-weighted centroids of areas i and j, that is, $\Sigma_{ij}(\phi) = f(d_{ij}, \phi)$. A simple one-parameter model that we consider in Section 4 has correlations given by $f(d_{ij}, \phi) = \exp(-d_{ij} \times \phi)$. One interpretation of ϕ follows from observing that $\log 2/\phi$ is the distance at which the spatial correlation drops to 0.5. Cressie [25] and Wackernagel [26] contain a discussion of more general forms for the correlation (including anisotropic possibilities). See also Richardson *et al.* [27] for comparative inference using different forms of $\Sigma_{ij}(\phi)$. An early example of a joint formulation based on distance (in an ecological setting) was given by Cook and Pocock [28].

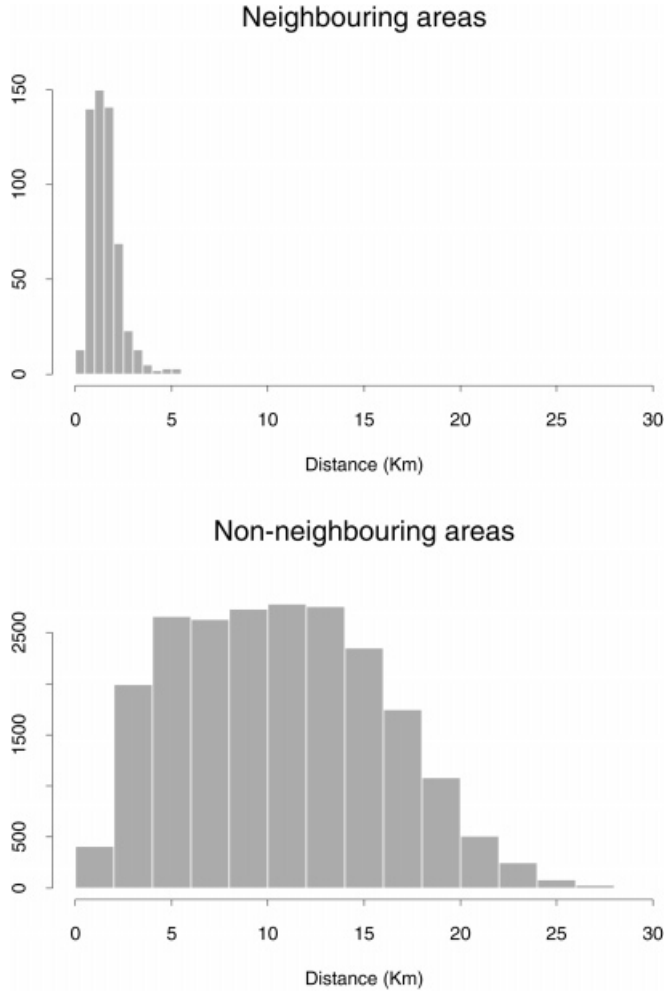

Figure 2. Histogram of between ward centroid distances for adjacent wards (top) and non-adjacent wards (bottom). Wards that share a common boundary are defined as neighbours.

In the conditional approach, the *intrinsic conditional autoregressive* (CAR) Markov random field prior suggested by Besag *et al.* [23] has commonly been used. This model is given by

$$U_i \mid U_j = u_j, j \neq i \sim \mathrm{N}\left(\bar{u}_i, \frac{\omega_u^2}{m_i}\right) \tag{6}$$

where $\bar{u}_i = \frac{1}{m_i} \sum_{j \in \partial i} u_j$, ∂i denotes the set of labels of the 'neighbours' of area i, and m_i is the number of such neighbours. In this model, we need to specify a neighbourhood scheme to reflect our beliefs about the spatial dependence structure between areas. The usual choice is for areas that share a common boundary to be defined as neighbours but this choice is less appealing if the study area contains areas of greatly varying size and shape. Figure 2 shows histograms of the distances between adjacent and non-adjacent wards for the larynx cancer data. For random variables on an

equally-spaced lattice, such histograms would not overlap, but here we see a degree of overlap. Hence a neighbourhood scheme based on distance might be appealing (see, for example, Best *et al.* [29]). An advantage of the distance-based models over the adjacency-based models is that missing areas (for example, with zero denominators) and discontinuities such as areas of water or woodland do not pose a difficulty for the former. In addition, models using neighbourhoods defined through adjacency also require the geographical boundaries to be available (as opposed to the centroids) and for older geographic units (for example, 1981 enumeration districts in the U.K.) this may pose a problem. More general neighbourhood structures have been suggested by Besag in the reply to the discussion of Besag *et al.* [23].

A difficulty with model (6) is the interpretation of the conditional variance ω_u^2. This quantity cannot be directly compared with σ_v^2 which is a marginal variance parameter, or with σ_u^2, the variance of the joint model. This difficulty of interpretation also makes prior choice more problematic. Similarly the nature of the spatial dependence in (6) is different from that implied by the joint model, since, by construction, the spatial correlation between neighbouring areas depends on the number of neighbours and only indirectly on the distance between neighbours. We also note that this model is non-stationary; this may be seen as an advantage since it is more flexible.

One great advantage of the conditional model is that it is very computationally efficient due to the conditional independencies that may be exploited in Markov chain Monte Carlo (MCMC) estimation approaches (Smith and Roberts [30]). In contrast, the joint model is computationally expensive since the parameters defining elements of $\Sigma(\phi)$ are treated as unknown random variables. Hence $\Sigma(\phi)^{-1}$ and its determinant are included in the Gaussian density of the joint posterior. As a result, a matrix inversion step is required at each iteration of the MCMC sampler for the joint model and hence this model requires far more computer time for implementation.

The estimation of the regression parameters β in model (5) is subject to the *ecological bias* that may result when group-level data are analysed and the inference is transferred to the level of the individual. In spatial epidemiological studies, the groups are geographical areas. This bias can occur for a number of reasons [31–35]. As usual in observational studies, bias can result from confounding, though the situation is more complex in ecological studies since confounders must be considered both between and within areas. If the exposure is non-constant across each area (which is almost certainly the case) then, unless the model is linear, the area-level model that results upon integration over the area does not correspond to the individual-level model; this leads to what Greenland [31] terms *pure specification bias*. Note that this form of bias can occur even in the absence of confounding. Mutual standardization of the response and explanatory variables must also be carried out to avoid bias [36]. For example, if the response is an age-standardized rate, then the exposure of interest should also be age-standardized in the same fashion (though unfortunately the data required to carry out this standardization will rarely be available). The effects of these various sources of bias are difficult to determine but within-area information on the exposure of interest and individual-level risk factors will be beneficial (see Lasserre *et al.* [37]).

At the third stage of the model, the Bayesian approach that we follow requires specification of prior distributions for the second-stage parameters β and ψ. Inference is carried out via MCMC simulation in which (dependent) samples are generated from the posterior distribution

$$p(U, V, \beta, \psi | y)$$

where $y = (y_1, \ldots, y_I)^{\mathrm{T}}$. For the analyses presented in this paper we use the WinBUGS software [38]. Two simulations were run for each model starting from different initial values; convergence

was checked by visual examination of trace plots and by calculation of the Brooks and Gelman [39] diagnostic.

3. NON-HIERARCHICAL ANALYSES

Chapter 4 of Breslow and Day [40] provides a detailed discussion of the fitting and interpretation of the multiplicative model (2). In particular they consider: the effects of marginal estimation of age effects compared to the (generally preferred) joint estimation; assessment of fit through Pearson's chi-squared and the deviance statistics; and embedding the multiplicative model within a broader family that includes the additive model.

We now consider how to assess the proportionality assumption (2). This may be examined informally via plots of $\log \hat{p}_{ij}$ versus $\log \hat{q}_j$, where $\hat{p}_{ij} = Y_{ij}/N_{ij}$ and $\hat{q}_j = \sum_i Y_{ij} / \sum_i N_{ij}$, in those areas containing sufficient data for reliable estimates. If (2) holds then these plots will produce a set of vertically shifted lines with slopes of one. Figure 3 shows this plot for 12 wards in our study region (selected among those with few empty strata). There is some indication that proportionality may not hold in the penultimate stratum (which corresponds to the 65–74 year olds) but interpretation is not straightforward since each point has a different precision associated with it.

A formal method of assessing proportionality may be carried out by fitting the Poisson model (1) with p_{ij} given by

$$\log p_{ij} = \mu + \alpha_i + q_j \tag{7}$$

and comparing it with the fit of the saturated model

$$\log p_{ij} = \mu + \alpha_i + q_j + \gamma_{ij} \tag{8}$$

using a likelihood ratio test (that is, via the deviance). As Table I shows, the data in each ward are very sparse and this casts doubt on the asymptotic chi-squared distribution of the deviance. In particular, with small expected numbers the observed value of the statistic may be far smaller than that predicted by the chi-squared distribution [41]. Hence we carried out a Monte Carlo test by simulating 999 replicate data sets from model (1) with p_{ij} given by (7), with values for μ, α_i and q_j fixed at their maximum likelihood estimates obtained from the observed data. Models (7) and (8) were then fitted to each of the observed and replicated data sets and the likelihood ratio test statistic evaluated. The rank order of the test statistic based on the observed data was used to derive the Monte Carlo significance level. This procedure led to a Monte Carlo p-value of 0.866 (based on 999 simulations), indicating no evidence that the simpler model should be rejected in favour of the saturated model, and hence that proportionality seems appropriate. We note that the main-effects model (7) gave a residual deviance of 789.4 on 848 degrees of freedom for the observed data. Using the asymptotic test would give $p = 0.996$. We note that the Monte Carlo test is slightly conservative since we have estimated parameters to carry out the simulation. On the basis of these results we have no reason to reject the assumption of proportionality.

As stated above, it is generally preferable to jointly estimate the age effects and the relative risks. However, it is often convenient to provide a set of age-specific rates q_j in order to reduce the number of data items by calculating the expected numbers in advance. Various possibilities are available including the use of an external set of rates, or internal standardization via $\hat{q}_j = Y_{+j}/N_{+j}$.

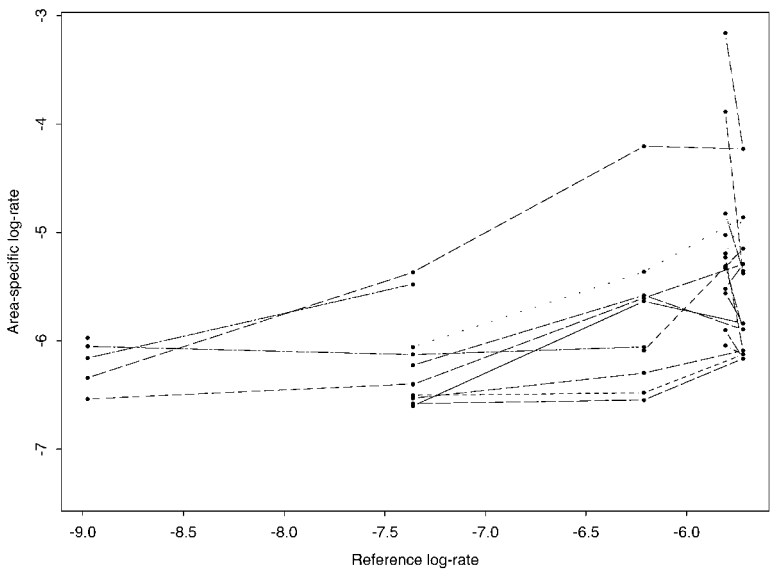

Figure 3. For selected wards, plot of $\log \hat{p}_{ij}$ versus $\log \hat{q}_j$, i indexes wards and j age groups. Parallel lines with slopes of one indicate that proportionality (equation (2)) may be appropriate.

In general, this approach will not produce identical estimates and standard deviations to the joint estimation of q_j and θ_i and care must be taken since separate estimation may remove some of the area effects. To assess whether there is any loss in carrying out the two-stage estimation, we fitted the model $\log p_{ij} = \mu + \alpha_i + q_j$, either with the parameters jointly estimated or with the q_j estimated separately and incorporated as offsets. Figure 4 displays age-specific rates calculated via these two approaches (and a Bayesian model that we describe in the next section). We see virtually no differences between the rates (and the standard errors are very similar also) and conclude that the two-stage approach using $\hat{q}_j$ is acceptable in this study.

Figure 5 plots the SMRs versus the Carstairs index. Each SMR has an associated 50 per cent confidence interval (this value was chosen to produce a clear plot) and a local scatter plot smoother was drawn through the set of three points (point estimate and endpoints of interval). We see that there is some indication of increasing relative risk with increasing deprivation for lower deprivation levels but for higher levels no such relationship can be seen. Figure 6 displays a map of the Carstairs index, which shows some suggestion of positive spatial correlation, that is, higher levels of deprivation in the south east and more affluent areas in the north of the study region.

Figure 5 highlights the presence of an outlying ward with an SMR of 7.77; this ward has 12 cases and an expected number of just 1.54. The cases in such a ward deserve special attention, in particular the recorded addresses should be checked to see if, for example, these cases have been assigned to the ward of the clinic they attended rather than to their residences. For the moment we retain these cases but for the geographic analyses described in the next section will scrutinize the effect of this outlier on inference.

In general it is of interest to see if there are any broad-scale geographic trends in the data and so we combine (3) with

$$\log \theta_i = \mu + \beta_e X_{ie} + \beta_n X_{in} \tag{9}$$

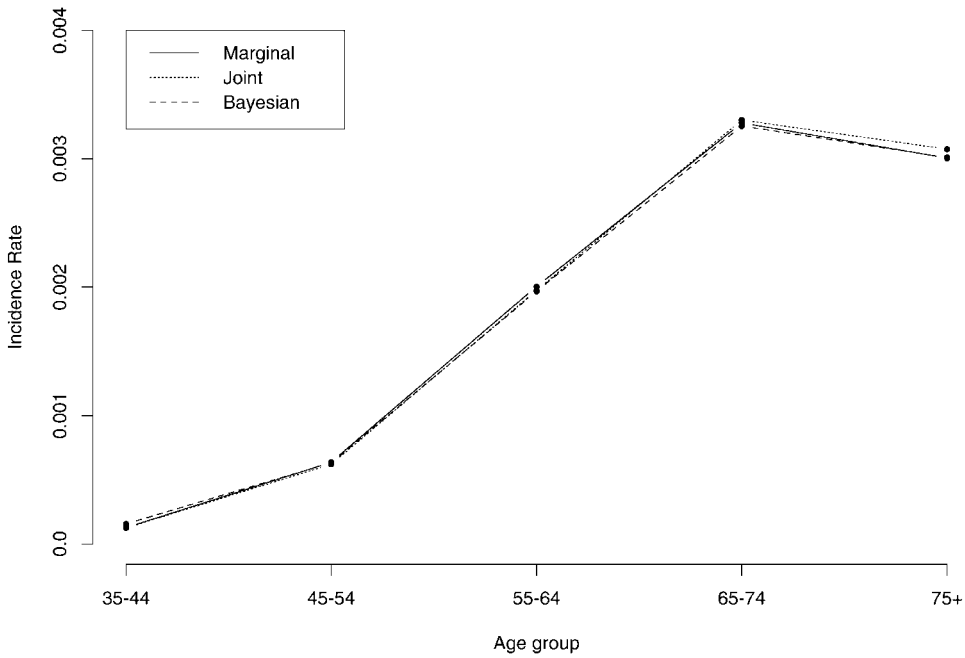

Figure 4. Age-specific rates from a variety of models.

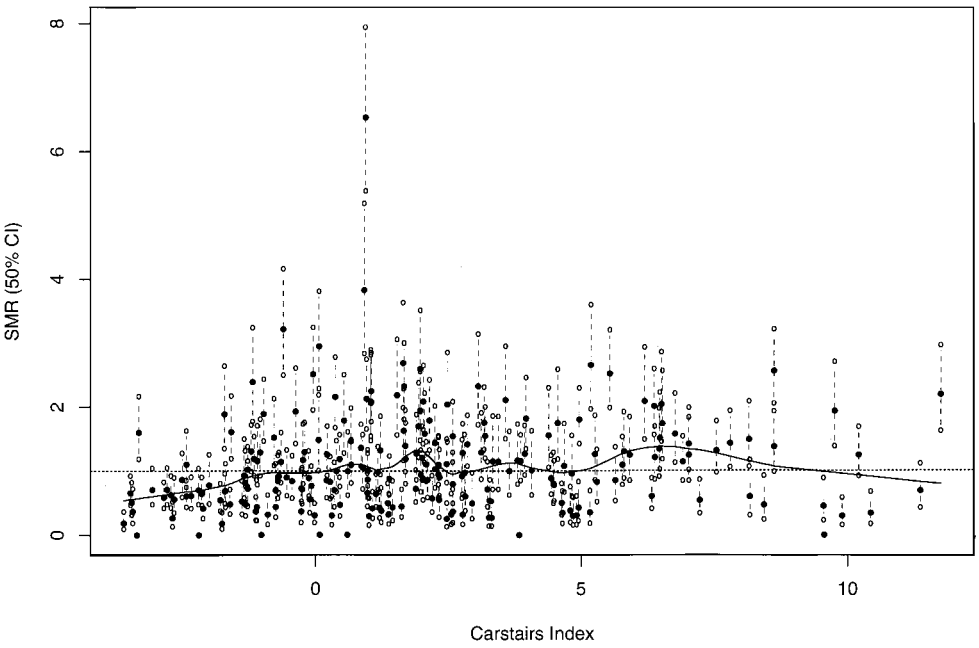

Figure 5. Estimated relative risks (SMRs), with 50 per cent confidence interval, plotted versus Carstairs deprivation score, with local smoother superimposed.

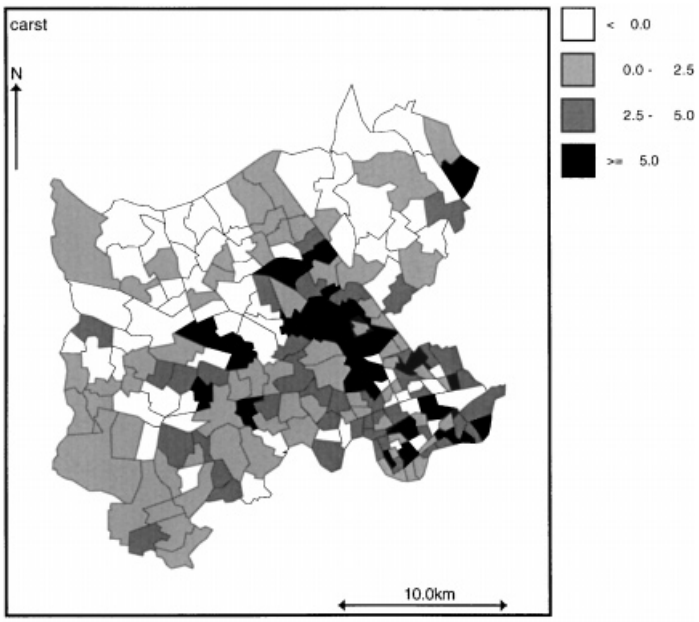

Figure 6. Map of Carstairs deprivation index; high values indicate greater deprivation.

where X_{ie} and X_{in} denote the eastings and northings of ward i, respectively. The fit of this model (assessed via a Monte Carlo deviance test, as before) was a substantial improvement over the model with no trend. We obtained $\exp(\hat{\beta}_e) = 1.02$ and $\exp(\hat{\beta}_n) = 0.90$ with both parameters being significantly different from one, and the latter highly so. This indicates an increase in relative risk of 2 per cent per kilometre in the eastings direction and a decrease of 10 per cent per kilometre in the northings direction. When the outlying ward is removed these estimates were virtually unchanged. Figure 7 shows the deviance residuals from model (9) plotted versus deprivation (with the outlier removed); again we see some suggestion of an association.

To assess the extent of extra-Poisson variability we may calculate the measure of overdispersion

$$\hat{\kappa} = \frac{\sum_i (Y_i - \hat{Y}_i)^2 / \hat{Y}_i}{I - k - 2}$$

(where we recall that k is the number of risk factors included in the log-linear model) corresponding to the quasi-likelihood approach in which $\mathrm{var}(Y) = \kappa \times E[Y]$ (for example, McCullagh and Nelder [42]). For model (9), that is, the model that assumes that the relative risks are only functions of eastings and northings, the measure of overdispersion is 4.98 with the outlying ward included, and 4.15 when this ward is excluded. Hence we conclude that, as expected, there is variability in the relative risks across the map. One aim of our analysis is to estimate the spatial and non-spatial components of this variability.

Various tests have been proposed to assess spatial autocorrelation in a set of disease rates. Walter [43] provides a critical review of a number of alternatives including the statistics due to Moran and Geary. Unfortunately the asymptotic moments of these statistics under the null hypothesis of no spatial dependence are incorrect when the constituent areas have different population sizes [44].

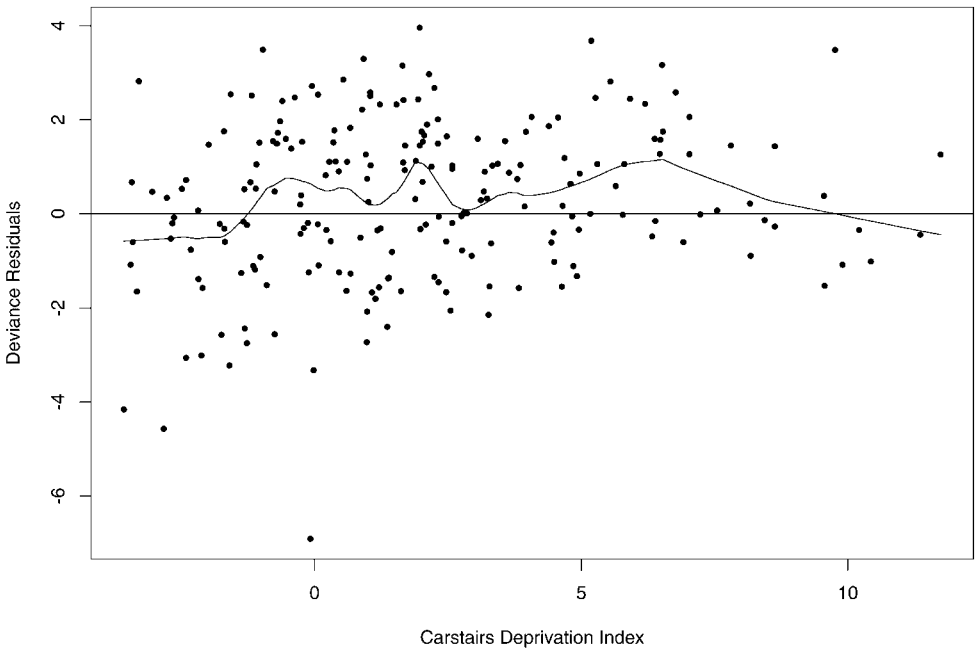

Figure 7. Deviance residuals versus Carstairs deprivation score, with local smoother superimposed.

This aspect may be fixed using a Monte Carlo test but another problem remains. Areas with small populations tend to be geographically close (for example, rural areas) and so will tend to display spatial dependence and the Moran and Geary statistics were not constructed with this aspect in mind. Alternatively, a number of authors (for example, Cressie and Chan [45] and Wakefield and Morris [46]) have considered examining the spatial dependence between sets of counts using the variogram. If we define $R(x)$ to be the value of the residual (here we consider deviance residuals, although Pearson residuals produce virtually identical results) at location x (that is, the population-weighted centroid of the area), and $d = |x - x'|$ to be the distance between x and x', we may then examine the isotropic spatial dependence at a distance d via the semi-variogram

$$\gamma(d) = \frac{1}{2}\mathrm{var}\{R(x) - R(x')\} \tag{10}$$

In practice, a discrete number of bins are considered and each of the observed distances is assigned to the bin that it corresponds most closely to. It is appropriate to use the variogram when the stochastic process under examination follows intrinsic stationarity [25]. In this case the mean is constant (with, in particular, no trend), and the right-hand side of (10) depends only on the vector difference between x and x' (and in particular the variance at any point is constant). As d becomes large, $\gamma(d)$ tends to a constant that is equal to the variance for a stationary process. The deviance residuals have approximately constant variance and to remove the trend we assume model (9). Figure 8 shows the variogram and we note that $\gamma(d)$ appears approximately constant at all distances (and equal to 1 which is the approximate variance of the standardised residuals). This indicates that the spatial dependence is not operating at large distances; at short distances there are fewer

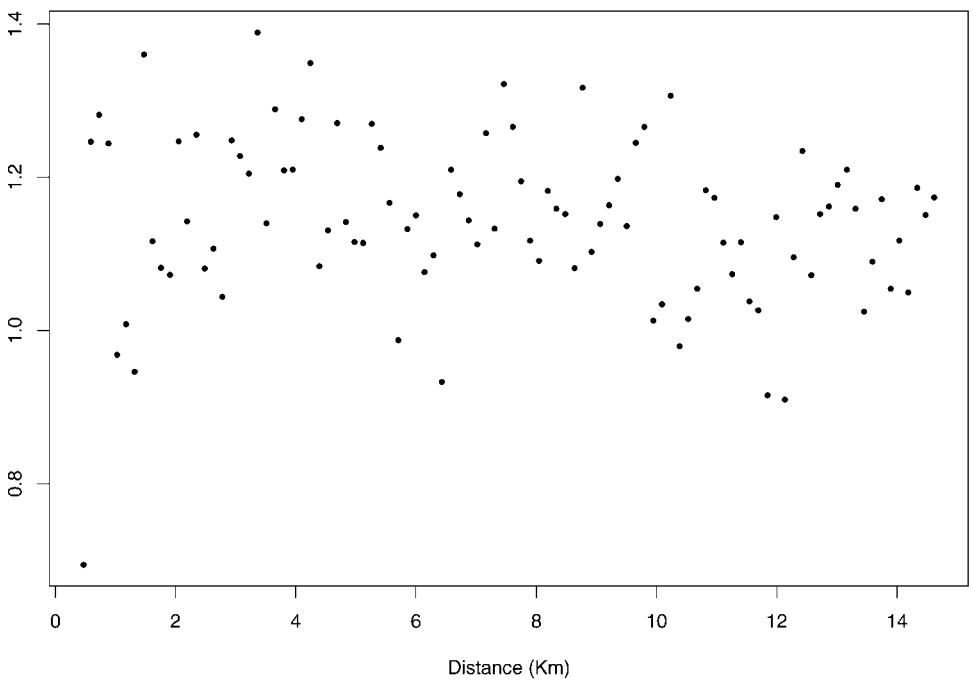

Figure 8. Variogram of deviance residuals from log-linear model in eastings and northings.

points from which the variogram may be estimated however (as indicated by Figure 8). We also constructed a variogram with the outlying ward excluded and found little difference.

Hence at this stage we have observed extra-Poisson variation, little broad scale spatial dependence, an indication of some relationship between risk and deprivation, and an outlying ward. We have also established that proportionality appears adequate and that little is lost by reducing the dimensionality of the data via the estimation of q_j in advance.

4. HIERARCHICAL ANALYSES

4.1. Effect of outlying area

As pointed out in Section 3, one of the wards produces an outlying relative risk and in this section we investigate the effect this outlier has on inference. We take model (3) at the first stage and assume, initially, that

$$\log \theta_i = \mu + V_i \tag{11}$$

with $V_i \sim_{\text{i.i.d.}} \text{N}(0, \sigma_v^2)$; this will be referred to as the 'heterogeneity only' model. An alternative distribution for the heterogeneity random effects is the Student's t-distribution with mean zero, scale parameter σ_v' and υ degrees of freedom, that is $V_i \sim_{\text{i.i.d.}} \text{St}_\upsilon(0, \sigma_v'^2)$; here we choose $\upsilon = 3$. The variance of a Student's t random variable, σ_v^2, is related to the scale parameter, $\sigma_v'^2$, via

Table II. Sensitivity of inference 'with' and 'without' the outlying ward, all parameters are posterior means. The 'Normal' and 'Student's t' models have $\log\theta_i = \mu + V_i$, with $V_i \sim_{\text{i.i.d.}} \text{N}(0, \sigma_v^2)$ and $V_i \sim_{\text{i.i.d.}} \text{St}_\upsilon(0, \sigma_v'^2)$, respectively. The 'Spatial' model has $\log\theta_i = \mu + V_i + U_i$, with $V_i \sim_{\text{i.i.d.}} \text{N}(0, \sigma_v^2)$ and an intrinsic CAR prior for U_i. The RR ratio refers to the ratio of the 95th and 5th percentiles of the distribution of the relative risks and θ_{outlier} the estimate of the ward containing the outlying data.

	Normal		Student's t		Spatial	
	With	Without	With	Without	With	Without
μ	−0.03	−0.02	−0.03	−0.02	0.00	−0.02
σ_v	0.28	0.07	0.23	0.09	0.11	0.04
ω_u	–	–	–	–	0.29	0.23
RR ratio	2.6	1.3	1.9	1.3	2.6	2.0
θ_{outlier}	2.1	0.99	6.2	0.99	2.0	1.3

$\sigma_v^2 = \sigma_v'^2 \times \upsilon/(\upsilon - 2)$. The advantage of the Student's t distribution is that it has heavier tails than the normal alternative and so inference is more robust to outlying areas. We also used model (11) with the addition of an intrinsic CAR random effect U_i (with U_i and V_i both assigned normal distributions, and the overall intercept term, μ, removed from the model). This will be referred to as the 'convolution model'. Table II summarizes inference for each of the analyses, with and without the outlier. In the analyses that excluded the outlying ward we included the number of cases for this ward as an unknown parameter. This is not necessary for the non-spatial models but it easily accounts for the addition of spatial random effects with the intrinsic CAR model. We note that if the joint model were used there would also be no need for the missing number of cases to be included as an unknown parameter, although this could still be done if inference concerning the true relative risk in the outlying area is of interest.

As expected, we see that the outlying area has a large influence on the results. In the heterogeneity only models the standard deviation of the log relative risks is four times greater when the distribution of the random effects is normal, and two and a half times greater with the Student's t distribution. This drastic change in σ_v is reflected in the ratio of the relative risks which is also greatly reduced when the outlying ward is removed. The Student's t distribution accommodates the outlying ward to some extent but sensitivity remains. When the outlier is excluded the relative risk of the missing area is estimated, as expected, as approximately 1 under the non-spatial models. Under the spatial models the relative risk is estimated as 1.3 because of the local smoothing that has been carried out. The top row of Figure 9 shows how the posterior means of the θ_i are related to the SMRs. We see that in each of the plots there is a large amount of shrinkage, particularly when the outlier is removed.

As a simple exploratory tool, we propose to measure the level of shrinkage in the heterogeneity only model via the following method. First consider the situation in which we have $Z \sim \text{N}(\phi, \sigma'^2)$ with σ'^2 known and the prior for ϕ given by $\phi \sim \text{N}(\mu, \sigma^2)$. In this case we have

$$E[\phi|Z] = w\mu + (1 - w)Z$$

where $w = \sigma'^2/(\sigma'^2 + \sigma^2)$, that is, the posterior mean is a weighted combination of the prior mean and the observed data. The weight $0 < w < 1$ can be interpreted as the degree of *shrinkage* to the prior; values close to 1 indicate high shrinkage. If we approximate the (first-stage) distribution

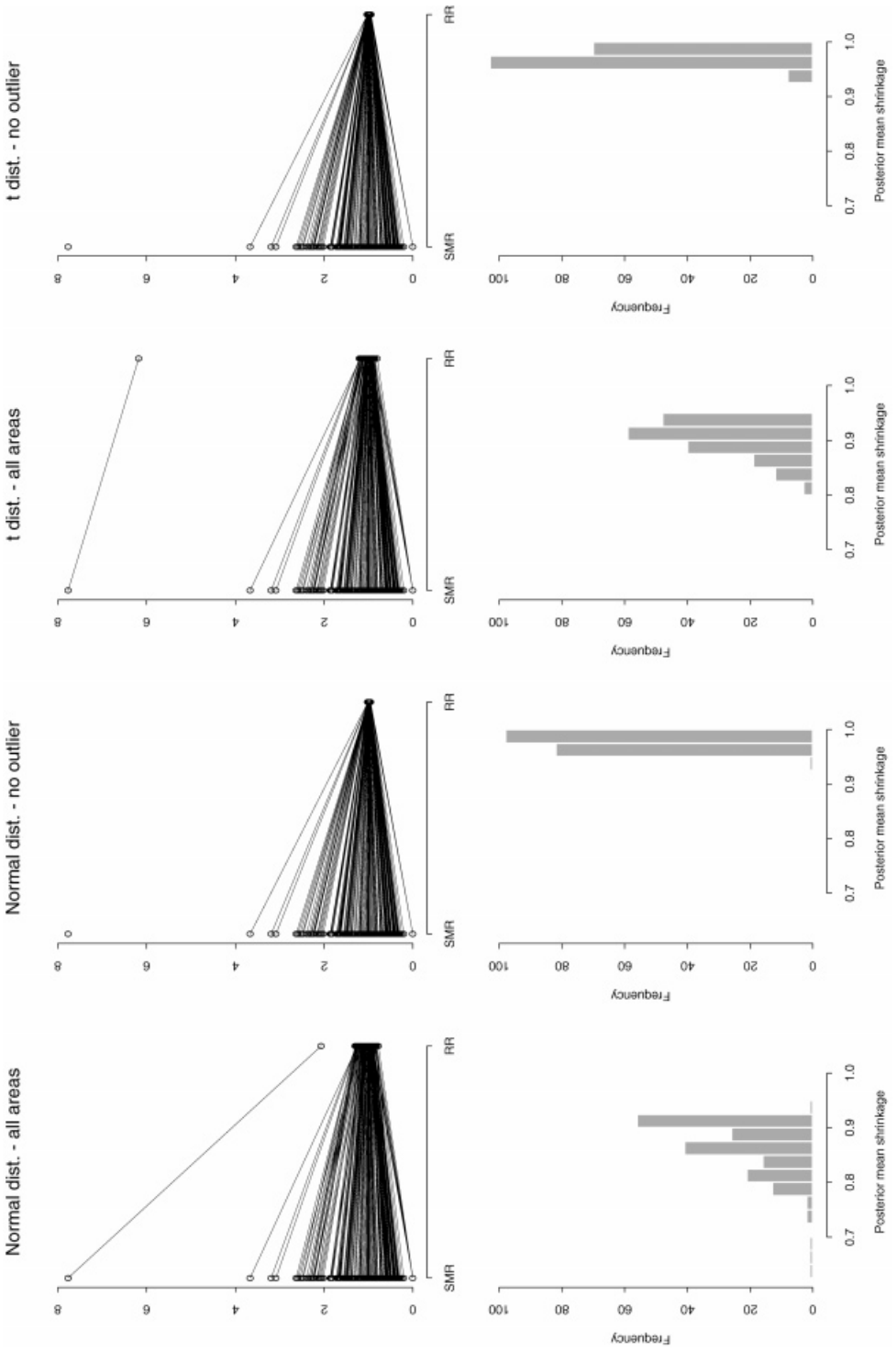

Figure 9. Measures of shrinkage under a variety of models and with and without the outlying ward. The top row shows the SMRs paired with the posterior means of each area and the bottom row the distribution of the weights given by (13).

of $\log \mathrm{SMR}_i$ by $\mathrm{N}(\log \theta_i, \sigma_i^2)$ then, by analogy, we can write

$$E[\log \theta_i | Y] = w_i E[\mu | Y] + (1 - w_i) \log \mathrm{SMR}_i \tag{12}$$

taking $\mathrm{SMR}_i = (Y_i + 0.5)/E_i$ if $Y_i = 0$ (for a justification for the addition of 0.5 see Clayton and Kaldor [22]). Re-arrangement of (12) gives

$$w_i = \frac{E[\log\ \theta_i | Y] - \log \mathrm{SMR}_i}{E[\mu | Y] - \log \mathrm{SMR}_i} \tag{13}$$

From (12) and (13) we see that a weight close to one indicates that a large amount of shrinkage to the overall mean has occurred. In the bottom row of Figure 9 we plot histograms of the weights w_i and see that, when the outlier is removed, this distribution is far more concentrated near 1. The high shrinkage is due to the small expected numbers in most wards, leading to large uncertainty in the SMRs and hence low weight given to the observed relative risk in each area.

This example shows that the data should be carefully examined for outliers as inference can be highly sensitive to their presence. In general, the cases that lead to the outlier should be investigated to determine whether, for example, the correct residential addresses were recorded or whether duplicates are present in the database. In the remaining analyses we exclude the outlier and treat the area as missing.

We note that little work has been carried out in the area of diagnostics (including outlier detection) in hierarchical models; see, however, the papers and accompanying discussion of Hodges [47] and Langford and Lewis [48].

4.2. Ecological association with deprivation

We now examine the relationship between risk and deprivation (as measured by the Carstairs index). A simple approach is to assume a log-linear relationship as in (5). The smooth lines in Figure 10 show estimates of the relative risk $\exp(\beta)$ corresponding to a unit change in the Carstairs index for each of the heterogeneity only and convolution models. We see that the relative risk is greatly reduced when the spatial component is included in the model. This is because the Carstairs index has spatial structure (Figure 6) and hence there is confounding between the index and the unmeasured risk factors that are being picked up by the spatial random effects. This supports the interpretation given by Besag *et al.* [23] that the U_i's are surrogates for unknown or unobserved spatially structured covariates, and that the ultimate goal is to identify and include sufficient covariate information so as to eliminate the need for spatial surrogates.

The above approach, though simple, may be too restrictive if the association is not of log-linear form. As an alternative we can discretize the deprivation index and treat the resultant variable as a factor. The coarseness of the discretization is a trade-off between the sparsness of the data and the flexibility produced. We choose to split the index into deciles but then place a first-order smoothing prior on the coefficients. This model is flexible but smooths out abrupt changes which is consistent with our prior beliefs of a smooth relationship. Specifically, if β_m represents the effect of level m of deprivation on the log relative risk, $m = 1, \ldots, 10$, we have

$$\beta_m | \beta_{m'},\ m' \neq m \sim \mathrm{N}(\beta_{m-1}, \lambda^2), \quad \text{for } d = 2, \ldots, 9$$

with β_1 assigned a flat prior. The parameter λ^2 is related to the variability in the factor levels β_m. Figure 10 shows the resultant relationship, via the posterior means of $\exp(\beta_m)$. We see that, as in

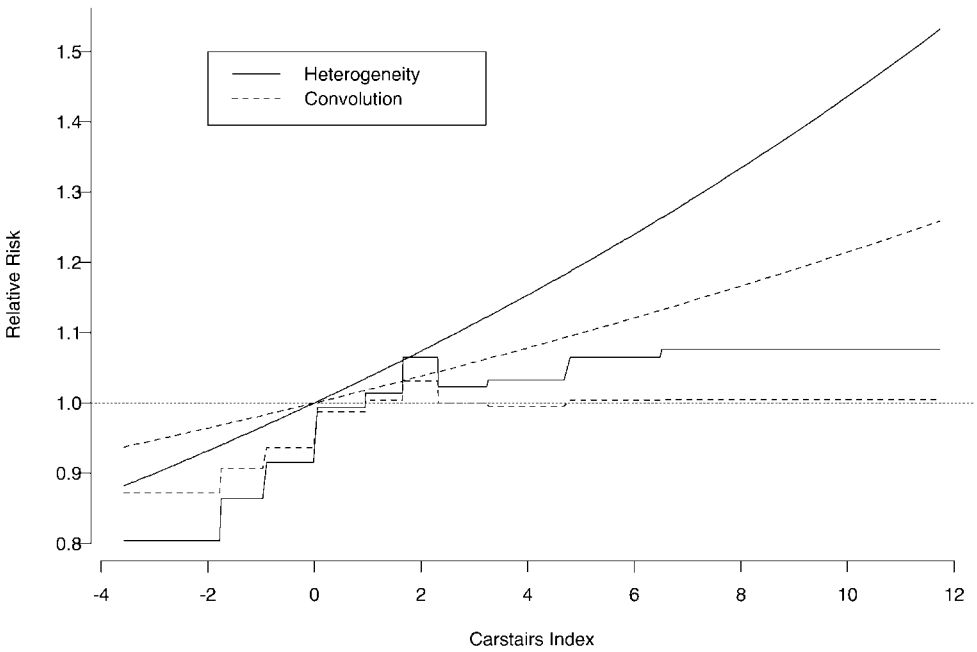

Figure 10. Association between relative risk and the Carstairs deprivation index using a log-linear regression and a first-order smoothing prior on the factor coefficients, for the heterogeneity and convolution second-stage models.

Figure 7, initially the relative risk increases with increasing deprivation but then flattens out at a deprivation score of around 4. Again we see that the relationship is attenuated under the spatial model. Hence we observe different behaviour to the simple log-linear model which demonstrates the advantage of the discrete approach from a descriptive point of view. For the sake of parsimony, a log-linear model with a threshold may provide a sensible compromise.

4.3. Prior sensitivity

A number of papers [29, 35, 49] have considered hyperprior specification and prior sensitivity with respect to the variance components σ_v^2, σ_u^2 and ω_u^2. We let $\mathrm{Ga}(a, b)$ denote the gamma prior with mean a/b and variance a/b^2 and examine three separate priors for the inverse of each variance component: (i) the choice $a = b = 0.001$ that has often been used as a default; (ii) the choice $a = 0.5$, $b = 0.0005$ that was suggested by Kelsall and Wakefield [50] to provide a plausible range for relative risks, and (iii) the choice $a = 10, b = 0.35$ which has been previously suggested by Bernardinelli *et al.* [49]. This latter choice corresponds to a strong prior belief that the ratio of the 5 per cent and 95 per cent relative risks in the heterogeneity only model is approximately 1.85. Under this model, for small σ_v^2, σ_v is approximately equal to the standard deviation of the relative risks. We also consider a fourth choice of prior, which is the uniform prior on the interval (0,1) for the variance parameters. On the interval (0,1), this prior shows similar behaviour to the inverse exponential prior proposed for the variance components of model (5) by Besag *et al.* [23]; however, their prior is improper and cannot be fitted using the WinBUGS software.

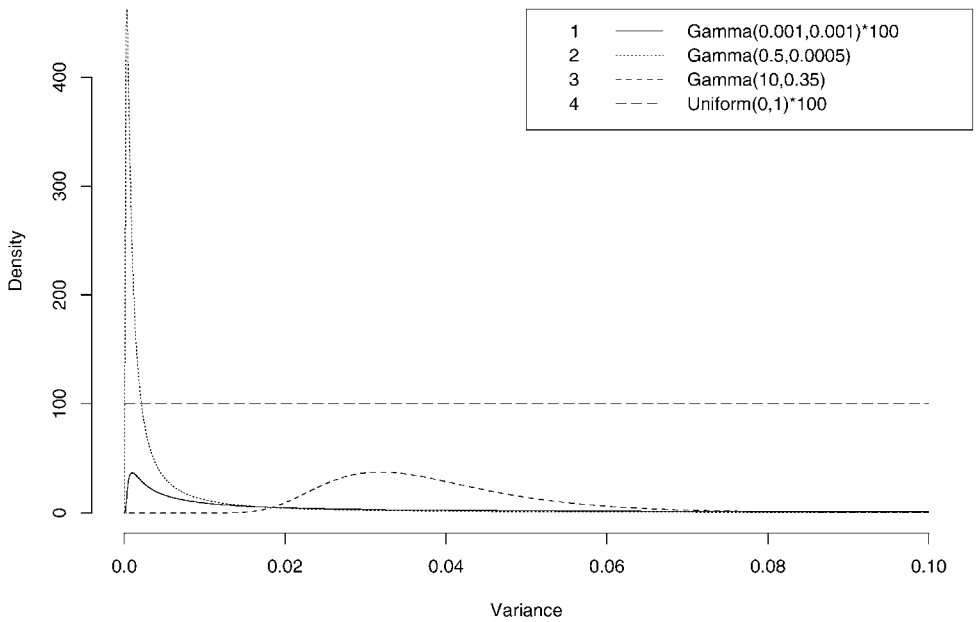

Figure 11. Comparison of four hyperprior distributions, shown on the scale of the random effects variance.

Figure 11 shows the four priors on the variance scale. We see that that the choice $a = 10$, $b = 0.35$ for the gamma distribution does not place much prior probability on small variances. Although the choice $a = b = 0.001$ is very flat, it also has little mass for very small variances.

For the joint model, the parameter ϕ that controls the strength of the correlations (via $\exp(-d\phi)$), was assigned a uniform prior on the range (l, u). The endpoints were chosen so that, if $d_{\max}$ denotes the maximum distance between centroids in the study region, the correlations at distances $d_{\max}/2$ and $d_{\max}/20$ were 0.01 which corresponds to $l = 0.16$ and $u = 1.6$. This prior acknowledges therefore that we cannot detect very short or very broad-scale correlations. The latter is because of non-identifiability between a non-zero log relative risk surface with no spatial dependence and a surface with zero mean and high spatial dependence.

Table III summarizes the results under the four priors, for each of the three models. We see that the standard deviation of the non-spatial random effects is very sensitive to the choice of prior with the Ga(0.5,0.0005) prior producing the smallest values and the U(0,1) prior the largest. This is consistent with Figure 11 since large values of the variance (and hence standard deviation) are more probable under the latter prior. The sensitivity is not evident for the spatial standard deviations since σ_u is not small and hence is consistent with all four priors. The relative risk ratios reflect the sensitivity to the non-spatial component. This provides further evidence that great care must be taken when prior distributions are assigned. The posterior mean estimates of ϕ imply that the correlations drop to 0.5 for areas that are at a distance between 1.8 km ($\phi = 0.39$) to 2 km ($\phi = 0.34$) apart. Hence the spatial dependence is acting at a relatively small scale (which is consistent with Figure 8). The relative sizes of σ_v and σ_u show that the spatial component is dominant.

Table III. Posterior mean summaries under three second-stage models and four third-stage hyperpriors. For the convolution prior σ_u represents the empirical standard deviation of the U_i, and so is comparable with σ_v.

Model	Prior	σ_v	σ_u	RR ratio	ϕ
Heterogeneity	Ga(0.001,0.001)	0.14	–	1.66	–
	Ga(0.5,0.0005)	0.07	–	1.29	–
	Ga(10,0.35)	0.19	–	1.89	–
	Un(0,1)	0.24	–	2.21	–
Convolution	Ga(0.001,0.001)	0.10	0.22	2.29	–
	Ga(0.5,0.0005)	0.04	0.21	2.03	–
	Ga(10,0.35)	0.18	0.20	2.43	–
	Un(0,1)	0.16	0.25	2.69	–
Joint	Ga(0.001,0.001)	0.14	0.24	2.59	0.39
	Ga(0.5,0.0005)	0.07	0.23	2.33	0.37
	Ga(10,0.35)	0.18	0.20	2.60	0.34
	Un(0,1)	0.14	0.27	2.87	0.36

Figure 12 shows the posterior means of the random effects plotted against each other for priors 1, 2 and 4 (the results for prior 3 are qualitatively similar to those for prior 1 and are not shown). The results of Table III are reinforced with the non-spatial random effects being sensitive to the prior choices and the spatial random effects being robust to this choice.

Figures 13–16 show the maps of the posterior means of the relative risks and the posterior probabilities that these relative risks exceed 1 under the heterogeneity-only model with Ga(0.5,0.0005) prior on σ_v^{-2}, the heterogeneity-only model with the U(0,1) prior on σ_v^2, and the convolution and joint models with Ga(0.5,0.0005) on each of the inverse variance components. We note that the heterogeneity-only map is virtually flat for the model using the Ga(0.5,0.0005) prior on σ_v^{-2}, while the convolution and joint maps, and to a lesser extent the heterogeneity-only map with U(0,1) prior on σ_v^2, display a region in the south-east with increased risks. This area of apparent high relative risk is consistent with the eastings and northings estimates obtained in Section 3. The convolution and joint models produce qualitively similar maps though there are differences in individual wards.

As a final check on prior sensitivity, we also considered replacing the Gaussian assumption for the conditional distribtion of the spatial random effects in the convolution model by a heavy-tailed double exponential distribution (see Besag *et al*. [23]). However, the results were virtually identical to those using the Gaussian model, so we do not report them here.

5. DISCUSSION

There are many unresolved issues in disease mapping. Careful use of techniques based on hierarchical models are an improvement over non-hierarchical models, however. In particular, disease maps produced using Bayesian smoothing methods are likely to be less visually misleading than their predecessors based on non-hierarchical approaches. The choice of prior distributions on the variance components σ_u^2, σ_v^2 and ω_u^2 in hierarchical models is clearly very important however. For ω_u^2 in particular this choice is non-intuitive because a conditional variance depends on the

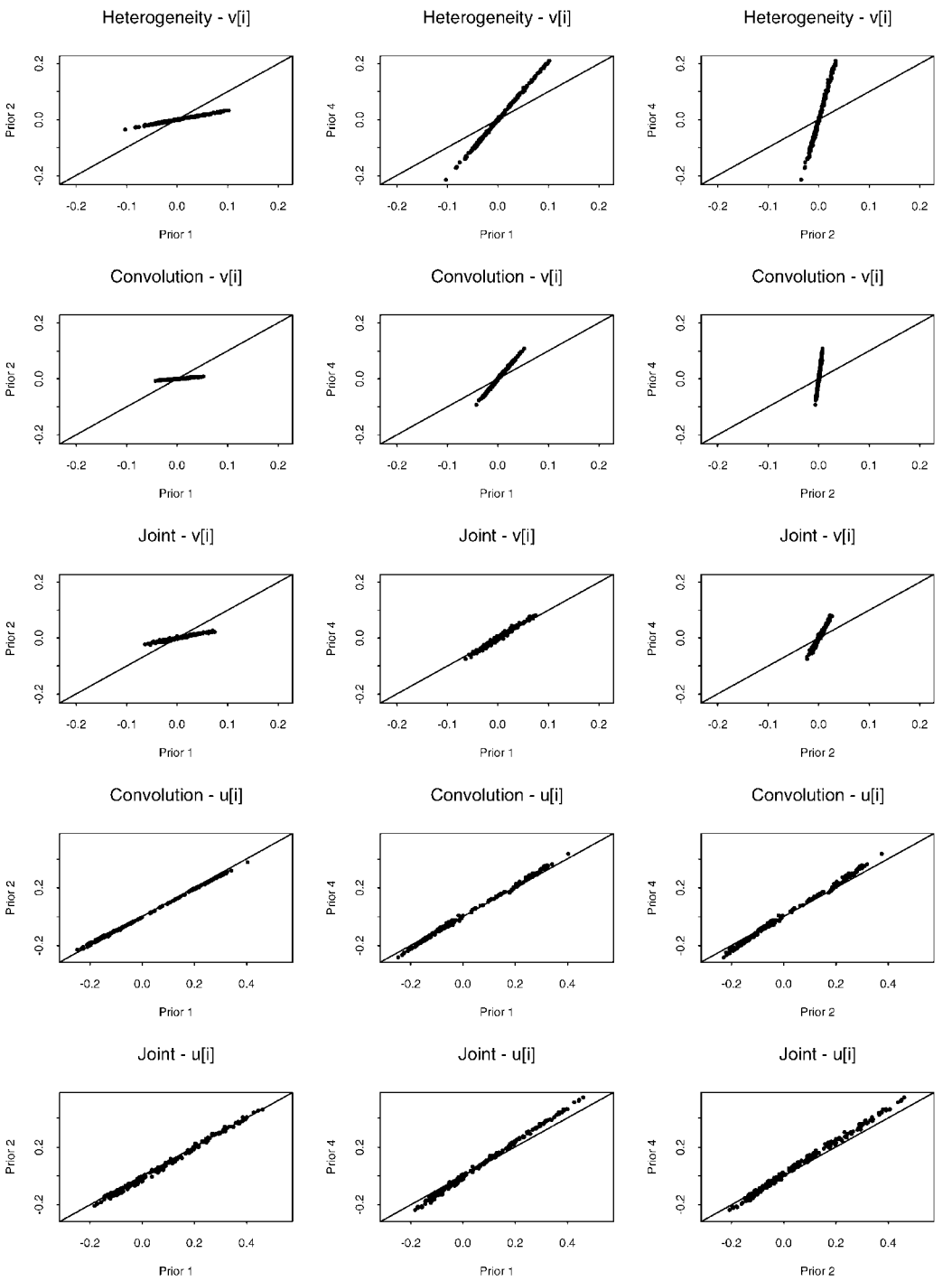

Figure 12. Association between spatial (U_i) and non-spatial (V_i) random effects under different models.

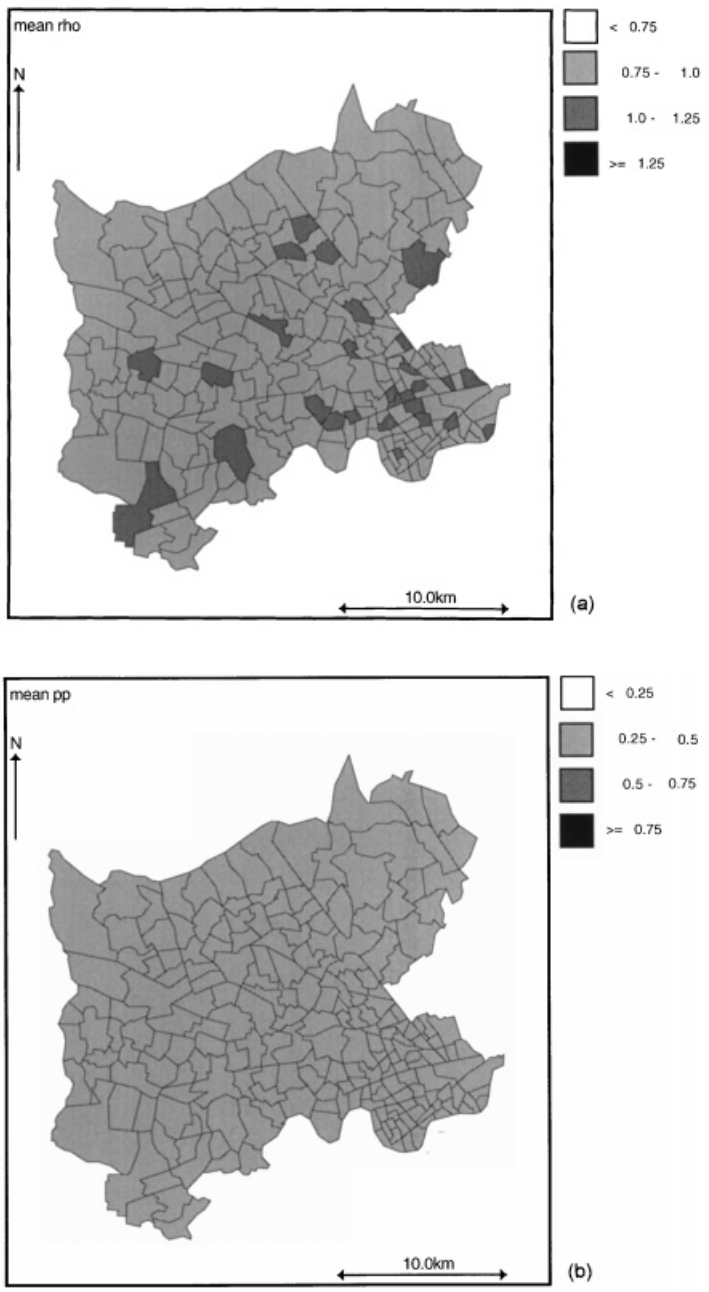

Figure 13. (a) Mapped posterior means $E[\theta_i|Y]$ and (b) posterior probabilities $\Pr(\theta_i > 1|Y)$, resulting from the heterogeneity-only model with a Ga(0.5,0.0005) prior on σ_v^{-2}.

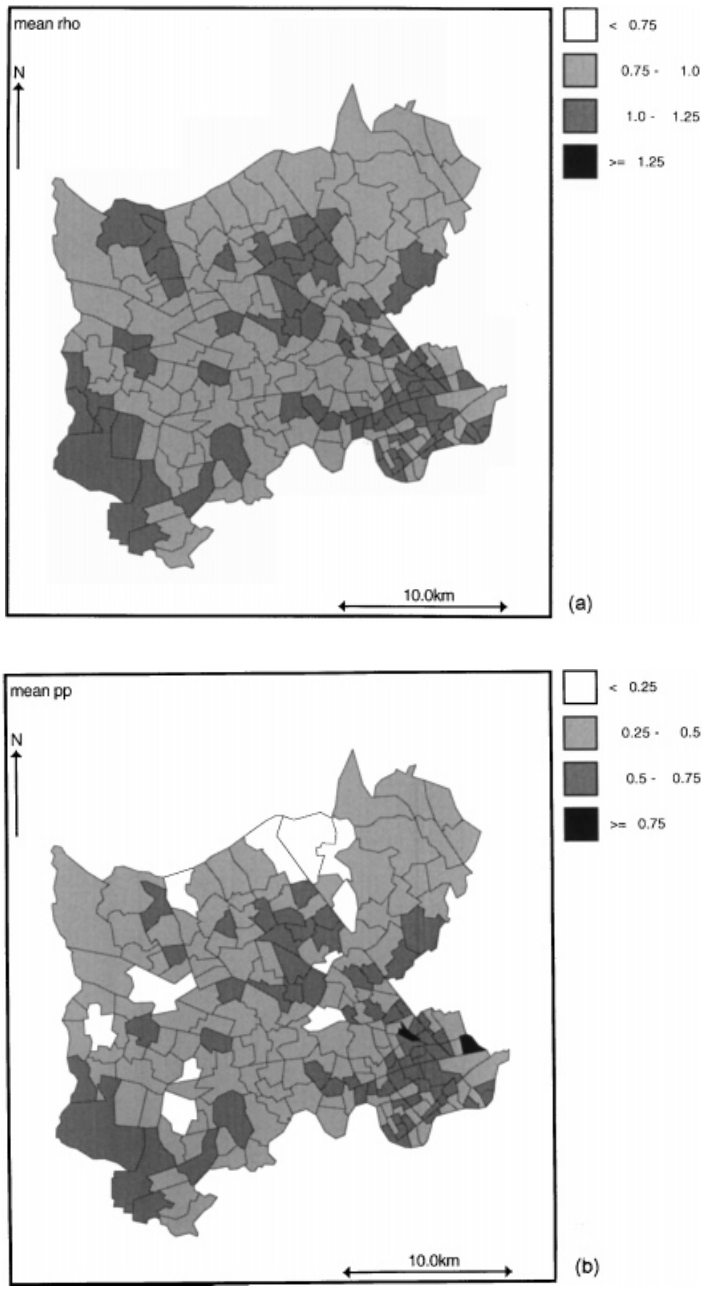

Figure 14. (*a*) Mapped posterior means $E[\theta_i|Y]$ and (*b*) posterior probabilities $\Pr(\theta_i > 1|Y)$, resulting from the heterogeneity only model with a U(0,1) prior on σ_v^2.

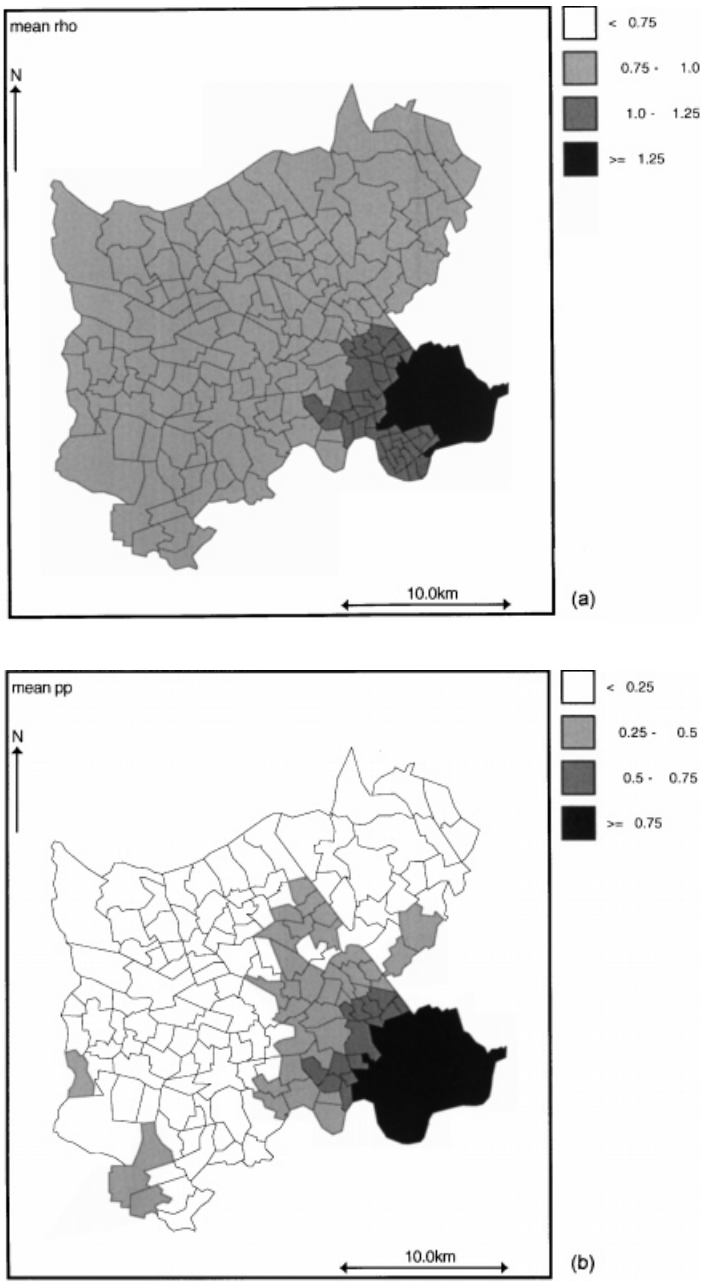

Figure 15. (*a*) Mapped posterior means $E[\theta_i|Y]$ and (*b*) posterior probabilities $\Pr(\theta_i > 1|Y)$, resulting from the convolution model with Ga(0.5,0.0005) priors on the inverse variance components.

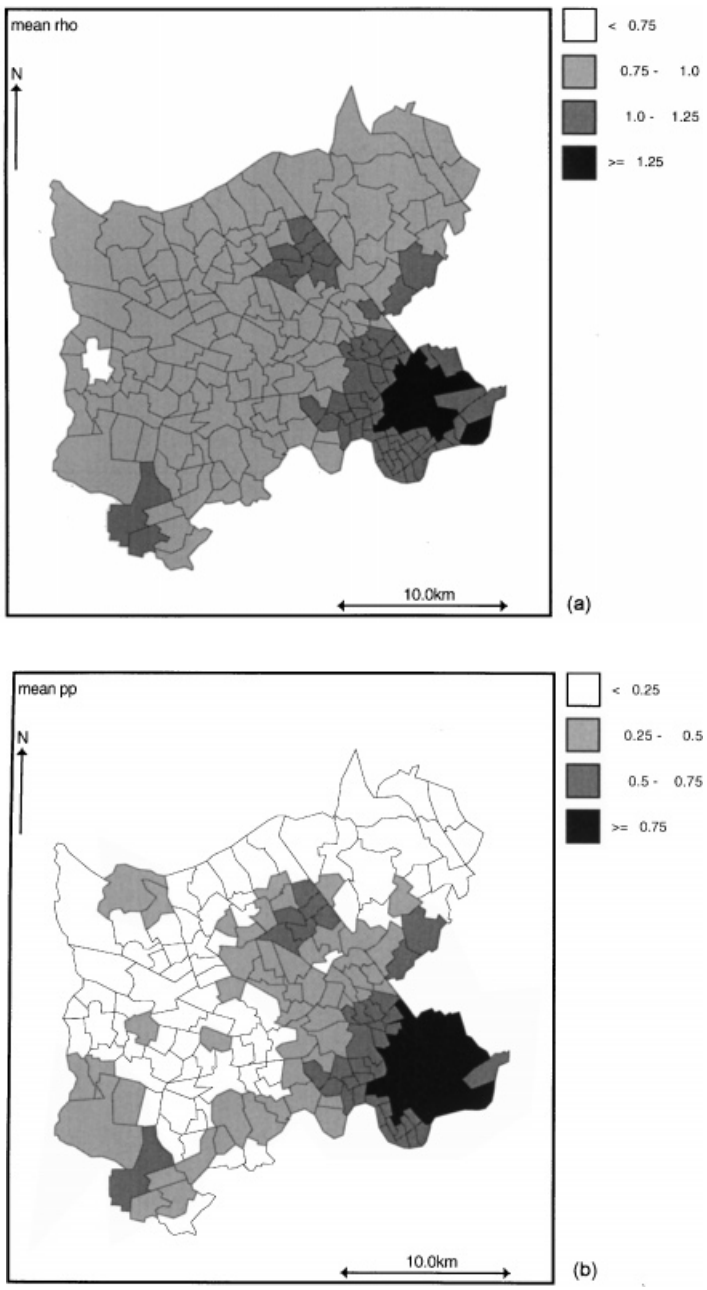

Figure 16. (*a*) Mapped posterior means $E[\theta_i|Y]$ and (*b*) posterior probabilities $\Pr(\theta_i > 1|Y)$, resulting from the joint model with Ga(0.5,0.0005) priors on the inverse variance components.

neighbourhood structure and in an epidemiological context this structure, in general, varies across the study area. Prior and posterior distributions may be graphically compared to informally assess the effect of the prior. We also note that when spatial and non-spatial random effects are introduced, one would expect negative correlation between the variance components and this is not usually acknowledged when priors are specified.

In general, the inclusion of spatially dependent random effects in disease mapping should be considered as a way of adjusting for unknown confounders. Hence sensitivity of the results to the specific model chosen for spatial dependence indicates that the results should be treated with great caution. Related to this issue are questions such as whether to include terms to model broad-scale trends, or whether a stationary or a non-stationary model is preferrable. For larger study areas in particular, we would expect the form of the spatial dependence to change across the region and hence some flexibility is required, though the information available to estimate the parameters of such a model will usually be sparse.

As an alternative to modelling the spatial dependence via a set of discrete random effects for each area, the underlying continous risk surface may be modelled. Such models are typically based on Poisson point processes; see, for example, Wakefield and Elliott [18], Lawson and Clark [51] and (for a non-epidemiological application) Wolpert and Ickstadt [24].

ACKNOWLEDGEMENTS

The authors would like to thank the Office for National Statistics who made the postcoded cancer data available for use. Population data came from the Estimating with Confidence project. This work is based on data provided with the support of the ESRC and JISC and uses census and boundary material which are copyright of the Crown, the Post Office and the ED-LINE Consortium. This work was partially funded by the Pan Thames Environmental R&D Programme, Project Reference Number 339 and grants from the ESRC (H519255036) and the EU BIOMED II program (PL96 3488). The data that are analysed form part of the Small Area Health Statistics Unit (SAHSU) database. This work was also supported, in part, by an equipment grant from The Wellcome Trust (0455051/Z/95/Z). We would also like to thank Andrew Thomas for providing assistance in the use of `WinBUGS` and Annie Mollié for helpful discussions.

REFERENCES

1. Clayton DG, Bernardinelli L, Montomoli C. Spatial correlation in ecological analysis. *International Journal of Epidemiology* 1993; **22**:1193–1202.
2. Parkin DM, Whelan SL, Ferlay J *et al. Cancer Incidence in Five Continents. Volume VII*. IARC: Lyon, 1997.
3. Coleman MP, Esteve J, Damiecki P *et al. Trends in Cancer Incidence and Mortality*. IARC: Lyon, 1993.
4. Raitiola HS, Pukander JS. Etiological factors of laryngeal cancer. *Acta Oto Laryngologica Supplement* 1997; **529**:215–217.
5. Dosemeci M, Gokmen I, Unsal M *et al.* Tobacco, alcohol use, and risks of laryngeal and lung cancer by subsite and histologic type in Turkey. *Cancer Causes and Control* 1997; **8**:729–737.
6. Tavani A, Negri E, Franceschi S, Barbone F, La Vecchia C. Attributable risk for laryngeal cancer in northern Italy. *Cancer Epidemiology, Biomarkers Prevention* 1994; **3**:121–125.
7. Sokic SI, Adanja BJ, Marinkovic JP, Vlajinac HD. Risk factors for laryngeal cancer. *European Journal of Epidemiology* 1995; **11**:431–433.
8. Rothman KJ, Cann CI, Fried MP. Carcinogenicity of dark liquor. *American Journal of Public Health* 1989; **79**:1516–1520.
9. Zheng W, Blot WJ, Shu XO *et al.* Diet and other risk factors for laryngeal cancer in Shanghai, China. *American Journal of Epidemiology* 1992; **136**:178–191.
10. Faggiano F, Lemma P, Costa G *et al.* Cancer mortality by educational level in Italy. *Cancer Causes and Control* 1995; **6**:311–320.
11. Esteve J, Riboli E, Pequignot G *et al.* Diet and cancers of the larynx and hypopharynx: the IARC multi-center study in southwestern Europe. *Cancer Causes and Control* 1996; **7**:240–252.

12. Guberan E, Usel M, Raymond L, Fioretta G. Mortality and incidence of cancer among a cohort of self employed butchers from Geneva and their wives. *British Journal of Industrial Medicine* 1993; **50**:1008–1016.
13. Goldberg P, Leclerc A, Luce D *et al.* Laryngeal and hypopharyngeal cancer and occupation: results of a case control-study. *Occupational and Environmental Medicine* 1997; **54**:477–482.
14. Elliott P, Hills M, Beresford J, Kleinschmidt I, Jolley D, Pattenden S, Rodrigues L, Westlake A, Rose G. Incidence of cancer of the larynx and lung near incinerators of waste solvents and oils in Great Britain. *Lancet* 1992; **339**:854–858.
15. Sans S, Elliott P, Kleinschmidt I, Shaddick G, Pattenden S, Walls P, Grundy C, Dolk H. Cancer incidence and mortality near the Baglan Bay petrochemical works, South Wales. *Occupational and Environmental Medicine* 1995; **52**:217–224.
16. Clayman GL, Stewart MG, Weber RS *et al.* Human papillomavirus in laryngeal and hypopharyngeal carcinomas. Relationship to survival. *Archives of Otolaryngology and Head and Neck Surgery* 1994; **120**:743–748.
17. Hoem JM. Statistical analysis of a multiplicative model and its application to the standardization of vital-rates—a review. *International Statistical Review* 1987; **55**:119–152.
18. Wakefield JC, Elliott P. Issues in the statistical analysis of small-area health data. *Statistics in Medicine* 1999; **18**:2377–2399.
19. Carstairs V, Morris R. *Deprivation and Health in Scotland.* Aberdeen University Press: Aberdeen, 1991.
20. Jolley D, Jarman B, Elliott P. Socio-economic confounding. *In Geographical and Environmental Epidemiology*: *Methods for Small-area Studies*, Elliott P, Cuzick J, English D, Stern R (eds). Oxford University Press: Oxford, 1992; 115–124.
21. Mollié A. Bayesian mapping of disease. *In* Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice.* Chapman and Hall: London, 1996.
22. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–682.
23. Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics* 1991; **43**:1–59.
24. Wolpert RL, Ickstadt K. Poisson/gamma random field models for spatial statistics. *Biometrika* 1998; **85**:251–267.
25. Cressie NAC. *Statistics for Spatial Data*, revised edn. Wiley: New York, 1993.
26. Wackernagel H. *Multivariate Geostatistics*, second completely revised edn. Springer-Verlag: Berlin, 1998.
27. Richardson S, Guihenneuc C, Lasserre V. Spatial linear models with autocorrelated error structure. *Statistician* 1992; **41**:539–557.
28. Cook DG, Pocock SJ, Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. *Biometrics* 1983; **39**:361–371.
29. Best NG, Waller LA, Thomas A, Conlon EM, Arnold R. Bayesian models for spatially correlated disease and exposure data. *In Bayesian Statistics 6*: *Proceedings of the Sixth Valencia Meeting on Bayesian Statistics*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford, 1999; 131–156.
30. Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 1993; **55**:3–23.
31. Greenland S. Divergent biases in ecological and individual-level studies. *Statistics in Medicine* 1992; **11**:1209–1223.
32. Greenland S, Robins J. Ecological studies-biases, misconceptions and counterexamples. *American Journal of Epidemiology* 1994; **139**:747–760.
33. Richardson, S. Statistical methods for geographical correlation studies. *In Geographical and Environmental Epidemiology*: *Methods for Small-area Studies*, Elliott P, Cuzick J, English D, Stern R (eds). Oxford University Press: Oxford, 1992; 181–204.
34. Morgenstern H. Ecological Study. *In Encyclopedia of Biostatistics*, *Volume 2*, Armitage P, Colton T (eds). Wiley: Chichester, 1998; 1255–1276.
35. Richardson S, Monfort C. Ecological correlation studies. In *Spatial Epidemiology*: *Methods and Applications*, Elliott P, Wakefield JC, Best NG, Briggs DB (eds). Oxford University Press: Oxford, 2000.
36. Rosenbaum PR, Rubin DB. Difficulties with regression analyses of age-adjusted rates. *Biometrics* 1984; **40**:437–443.
37. Lasserre V, Guihenneuc-Jouyaux C, Richardson S. Biases in ecological studies: utility of including within-area distribution of confounders. *Statistics in Medicine* 2000; **19**:45–59.
38. Spiegelhalter DJ, Thomas A, Best NG. *WinBUGS User Manual*, *version 1.1.1*. Cambridge: U.K., 1998 (available from `http://www.mrc-bsu.cam.ac.uk/bugs`).
39. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**:434–455.
40. Breslow NE, Day NE. *Statistical Methods in Cancer Research*, *volume II*: *The Design and Analysis of Cohort Studies*. IARC Scientific publications, Number 82, 1987.
41. Bithell JF, Dutton SJ, Neary NM, Vincent TJ. Controlling for socio-economic confounding using regression methods. *Journal of Epidemiology and Community Health Supplement 2* 1995; **49**:S15–S19.
42. McCullagh P, Nelder JA. *Generalised Linear Models*, 2nd edn. Chapman and Hall: London, 1989.
43. Walter SD. Assessing spatial patterns in disease rates. *Statistics in Medicine* 1993; **12**:1885–1894.

44. Besag J, Newell J. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society*, *Series A* 1991; **154**:143–155.
45. Cressie N, Chan NH. Spatial modelling of regional variables. *Journal of the American Statistical Association* 1989; **84**:393–401.
46. Wakefield JC, Morris SE. Spatial dependence and errors-in-variables in environmental epidemiology. *In Bayesian Statistics 6*: *Proceedings of the Sixth Valencia Meeting on Bayesian Statistics*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford, 1999; 657–684.
47. Hodges JS. Some algebra and diagnostics for hierarchical models (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1998; **60**:497–536.
48. Langford IH, Lewis T. Outliers in multilevel data (with discussion). *Journal of the Royal Statistical Society*, *Series A* 1998; **161**:121–160.
49. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine* 1995; **14**:2411–2431.
50. Kelsall JE, Wakefield JC. Discussion of 'Bayesian models for spatially correlated disease and exposure data', by Best NG, Waller LA, Thomas A, Conlon EM, Arnold R. *In Bayesian Statistics 6*: *Proceedings of the Sixth Valencia Meeting on Bayesian Statistics*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford, 1999; 151.
51. Lawson AB, Clark A. Markov chain Monte Carlo methods for putative sources of hazard and general clustering. *In Disease Mapping and Risk Assessment for Public Health*, Lawson AB, Boehning D, Lesaffre E, Biggeri A, Viel JF, Bertollini R (eds). Wiley: Chichester, 1999, 119–142.

